Novel mycoviruses discovered from metatranscriptomics survey of soybean phyllosphere phytobiomes

Shin-Yi Lee Marzano^{ab} and Leslie Domier^{ac}

^aDepartment of Crop Sciences, University of Illinois, ^cUnited States Department of Agriculture - Agricultural Research Service, Urbana, IL USA

^bCurrent address: Department of Biology and Microbiology, Department of Plant Science, South Dakota State University, Brookings, SD USA

Keywords: high-throughput sequencing metatranscriptome, phyllosphere, soybean, viromes

* Corresponding author at: Department of Crop Sciences, University of Illinois, United States Department of Agriculture - Agricultural Research Service, Urbana IL 61801 USA.

Fax: 217-333-5251.

E-mail address: leslie.domier@ars.usda.gov (L.L. Domier).

Abstract

Mycoviruses can be beneficial to plants in that they can debilitate pathogenic fungi thereby reducing the severity of associated plant diseases. Studies to date have focused primarily on culturable fungi that represent a fraction of natural fungal populations. The nonculturable fungi, however, can harbor diverse populations of mycoviruses that reduce plant disease or enhance resistance to abiotic stress. Metatranscriptome analysis of field-grown plant samples using high-throughput sequencing offers the possibility of unbiased detection and quantification of mycoviruses regardless of the culturability of their fungal hosts together with the complete associated microbial consortia. In this study, we describe the fungal viromes of the phyllosphere of production soybean fields in Illinois, USA by analyzing the metatranscriptomes of thousands of soybean leaf samples collected during the 2008 and 2009 growing seasons. The analyses identified 25 partial genome sequences that represented at least 22 mycovirus genomes, only one of which had been described previously. The novel mycovirus genomes showed similarity to 10 distinct lineages including the genera *Alphapartitivirus*, *Botybirnavirus*, *Endornavirus*, *Mitovirus*, *Mycoflexivirus*, *Ourmiavirus*, *Totivirus*, *Victorivirus*, family *Tombusviridae*, order *Mononegavirales*, and the recently proposed genus *Gemycircularvirus*. The present study adds to the wealth of mycoviruses associated with plant phytobiomes and establishes groundwork needed for further characterization of the viruses.

1. Introduction

Phytobiomes, the microbial communities associated with plants, include bacteria, fungi and viruses, and are increasingly recognized as important factors in plant health and productivity

(Hale et al., 2014). An expanding number of mycoviruses are being described that are directly or indirectly antagonistic to economically important fungal plant pathogens (Ghabrial et al., 2015), including mycoviruses that debilitate the causal agents of chestnut blight (Anagnostakis, 1982), Dutch elm disease (Brasier, 1983; Hong et al., 1999), Sclerotinia stem rot (Hu et al., 2014; Khalifa and Pearson, 2014a; Liu et al., 2014; Marzano et al., 2015; Yu et al., 2013), white root rot (Chiba et al., 2009; Kondo et al., 2013b) and Victoria blight of oats (Ghabrial et al., 2002). Yet, little is known about mycoviruses associated with natural fungal populations, pathogenic or not, that can be found in association with aerial portions of plants.

All debilitating mycoviruses have been isolated from cultured fungal isolates. However, Blackwell et al. (2011) estimated that less than 2% of the extant fungal species have been identified and that most fungi, like most bacteria, cannot be cultured by traditional methods. For example, many beneficial endophytic fungi (Rodriguez et al., 2009) and even common biotrophic plant pathogens like *Podosphaera xanthii* that causes powdery mildew (Perez-Garcia et al., 2009) and *Puccinia graminis* that causes wheat stem rust (Leonard and Szabo, 2005) are not amenable to *in vitro* culture. Therefore, analysis of communities *in situ* is required for some fungi and has the potential to identify higher order interactions of mycoviruses in disease suppression. Through these types of analyses, it may be possible to associate specific mycoviruses with enhanced plant vigor and reduced incidence of or sensitivity to fungal diseases. A study by Marquez et al. (2007) described a mycovirus that likely imparted thermal tolerance to a fungal endophyte and its plant hosts. In addition to the use of hypovirus-infected *Cryphonectria parasitica* for chestnut blight control (Nuss, 1992), the commercial application of fungal species *Trichoderma* and *Gliocladium* for suppression of plant disease (McSpadden-

Gardener and Fravel, 2002; Woo et al., 2014) provide precedences for manipulating fungal communities in the management of plant disease.

Recently, high-throughput sequencing has become an important tool in metagenomics because it does not require prior knowledge of specific genome sequences, has high sensitivity and is able to discover novel sequences (Mokili et al., 2012; Roossinck, 2015). High-throughput sequencing has been used for mycovirus discovery by analyzing double-stranded RNAs (Al Rwahnih et al., 2011; Coetzee et al., 2010; Roossinck et al., 2010), small interfering RNAs (Vainio et al., 2015), mRNA (Schoebel et al., 2014) and ribosomal RNA (rRNA)-depleted total RNAs (Chandler et al., 2015; Marvelli et al., 2014; Marzano et al., 2015). In addition, transcriptomes of fungi downloaded from publicly available databases provided evidence of the existence of negative-sense RNA viruses in fungi (Kondo et al., 2013a) before the first negative-sense RNA mycovirus was isolated (Liu et al., 2014). Analysis of total RNAs is capable of detecting non-polyadenylated viral RNAs and transcripts of DNA viruses, and consequently provides the most unbiased assessment of fungal viromes. Sequence-based categorization of viromes is a first step to gain a grand view of the mycoviruses that exist in both the endosphere and the interface between plants and the environment.

In previous studies, we assembled near complete genomes of four new mycoviruses from metatranscriptomics data from fungal cultures, which allowed the construction of a full-length cDNA clone and production of infectious synthetic transcripts for one of the mycoviruses (Marvelli et al., 2014; Marzano et al., 2015). Comparison of the new mycoviral genome sequences to other strains from widely dispersed geographic locations revealed inter- and intraspecific recombination events that provided insights into viral evolution (Marzano et al., 2015). In this study, we used a culture-independent metatranscriptomics approach to characterize

the mycovirus communities of natural fungal populations associated with field-grown soybean (*Glycine max* L. Merr) and identified near full-length sequences of putative mycoviral genomes. The identified mycoviruses can be exploited as natural genetic resources to further describe the core fungal virome that shapes natural tritrophic interactions among soybean plants, fungi and mycoviruses.

2. Materials and Methods

2.1. Plant materials and total RNA

In 2008 and 2009, soybean leaf samples were collected without regard to disease symptoms at growth stages R5 to R7 (Fehr et al., 1971) from commercial soybean fields in 36 and 9 counties respectively, in Illinois, USA (Hobbs et al., 2010). Leaf samples (over 3,000 in total) were combined into groups of about 90 leaves, total RNA was extracted using the Qiagen (Valencia, CA) RNeasy Plant Mini Kit for 150 mg of leaf tissue from each pool. RNA samples were treated with DNase I and evaluated for integrity by agarose gel electrophoresis (Hobbs et al., 2010).

2.2. High-throughput sequencing

Total RNA (approximately 1 µg) free of genomic DNA was submitted to the W. M. Keck Center, University of Illinois for processing. RNA samples were depleted of rRNAs with the Ribo-Zero Plant Kit (Epicentre, Madison, WI), used as templates to construct paired-end

libraries with a TrueSeq stranded RNA sample preparation kit (Illumina, San Diego, CA) and sequenced on an Illumina HiSeq2500. Separate libraries were constructed for each year, combined and sequenced in a single lane to produce 100 nt reads.

2.3. Sequence analysis

To assess the diversity of organisms associated with soybean leaves, sequence reads were aligned using DIAMOND (Buchfink et al., 2015) to a subset of the National Center for Biotechnology Information (NCBI) nonredundant (nr) amino acid sequence database, parsed and assigned to taxa using MEGAN5 (Huson et al., 2011). Sequence reads from both years were also combined and assembled *de novo* with Inchworm from Trinity (Grabherr et al., 2011) with a kmer of 25. Contigs with similarity to mycoviruses were identified using USEARCH (Edgar, 2010) and a custom database containing fungal, soybean and viral amino acid sequences from GenBank. Putative viral contigs from Inchworm assemblies were extended with the paired-read iterative contig extension (PRICE) program (Ruby et al., 2013) with parameters of 30 cycles, 100 nt for maximum length of sequence that fed into de Bruijn assembly, 30 nt for minimum overlap length for mini-assembly, 20 nt as the threshold sequence number for scaling overlaps for contig-edge assemblies, and minimum 80% identity for contig-edge assembly, 90% identity to the initial input contig to count as a match. The nucleotide sequences of all suspected mycovirus contigs were compared with the NCBI nr database using blastn (Altschul et al., 1990) to exclude misidentified sequences. Raw reads that aligned to the putative mycovirus contigs using Bowtie 2 (Langmead and Salzberg, 2012) were *de novo* assembled again using Sequencher 5.1 (Gene Codes Corp, Ann Arbor, MI, USA) using the default parameters of minimum overlap

of 20-nt and minimum overlap identity of 85%. The numbers of reads aligning to different target sequences were calculated using Bowtie (Langmead et al., 2009). Predicted amino acid sequences were aligned using MUSCLE (Edgar, 2004) and maximum-likelihood phylogenetic trees were constructed using MEGA 6.0 (Tamura et al., 2013). To confirm that sequences similar to fungal RNA viruses were not derived from sequences integrated into host genomes and that the two contigs similar to gemycircularvirus coding regions were present within the same circular template, DNA was extracted from pools of the 2009 samples using the DNeasy Plant Mini Kit (Qiagen), amplified with virus-specific primers (Supplementary Table 1) and analyzed by agarose gel electrophoresis. The amplicon produced by inverse primers specific for the gemycircularvirus were sequenced using an Applied Biosystems 3730xl automated sequencer. Splice-donor and acceptor sites in the predicted Rep-encoding primary transcript were predicted manually.

3. Results

3.1. Viromes associated with leaves of field-grown soybean plants

High-throughput sequencing of the RNA samples from field-grown soybean plants produced 2.7×10^8 paired-end reads from the two years. The predicted amino acid sequences of approximately 31% of the reads aligned to the subset of the NCBI nr database, among which 18% were assigned to viruses, 54.2% to plants (soybean), 13.7% to fungi, and 6.9% to bacteria in MEGAN analyses. Among the reads assigned to viruses, 99.6% were derived from soybean-infecting viruses (primarily alfalfa mosaic virus, bean pod mottle virus and soybean vein necrosis virus) and 0.4% were derived from sequences that showed similarity to previously described

mycoviruses. The four fungal genera assigned the most reads in the two years, in decreasing order of abundance, were *Fusarium*, *Pyrenophora*, *Parastagonospora* and *Leptosphaeria*.

Analysis of contigs assembled from metatranscriptome data identified 25 sequences with homology to previously described taxa of mycoviruses (Table 1). Based on phylogenetic relationships, the putative mycovirus genomes were predicted to be composed of positive (+) or negative-sense (-) single-stranded (ss) RNA, double-stranded (ds) RNA or ssDNA. The predicted genome coverage ranged from 29% to greater than 95%. Near complete genomes (>90% coverage) included putative members of the genera *Alphapartitivirus* (RNA1), *Mycoflexivirus*, and *Mitovirus* (four sequences), along with a ss(-) RNA virus and a ss(+) RNA virus and an isolate of *Coniothyrium minitans* RNA virus 1 (Table 1). Among these, the near complete genomes of the putative alphapartitivirus, one of the mitoviruses and the ss(-)RNA virus were assembled from less than 1000 raw reads, indicating the sensitivity of the approach. The metatranscriptome from soybean leaves collected in 2008 contained more mycovirus-like sequences than those from 2009, possibly because the samples were collected later in the growing season than in 2008. All of the putative mycovirus genomes were detected in both years except for the ssDNA virus that was detected only in 2009. The sequence designated soybean leaf-associated mycoflexivirus 1 (SlMyfV) had the highest number of aligned reads and more than ten times as many raw reads as the largest putative ss(-)RNA virus, which had an assembled sequence of comparable size (Table 1). Attempts to detect the putative RNA viruses in DNA extracted from the 2009 field samples by polymerase chain reaction (PCR) did not produce amplicons, indicating that the sequences likely were not derived from virus segments integrated into host genomes (data not shown).

3.2. Putative flexivirus genome

One large contig SlaMyfV (8,327 nt; KT598226) contained one large open reading frame (ORF 1) and three small downstream ORFs (Fig. 1A). The contig had the highest number of aligning reads in 2008. The predicted amino acid sequence of ORF1 contained methyltransferase, helicase and RNA-dependent RNA polymerase (RdRp) domains and was most similar to the replicase of Botrytis virus F (BotVF; 6,827 nt) (Howitt et al., 2001), the type member of the genus *Mycoflexivirus* in the family *Gammapflexiviridae* and order *Tymovirales* (Adams et al., 2012d). However, the organization of the predicted ORFs (one large 5'-proximal ORF and multiple small 3'-proximal ORFs) was more similar to viruses in the *Alphaflexiviridae* and *Betaflexiviridae* (Adams et al., 2012a; Adams et al., 2012b) than those in the *Gammapflexiviridae* (Adams et al., 2012d). Phylogenetic analysis of the putative replicase amino acid sequence positioned the virus genome between the *Gammapflexiviridae* and *Tymoviridae* families (Fig. 1B). At a maximum of 43% amino acid sequence identity to BotVF, the sequence represented a novel virus genome. The predicted amino acid sequences of the downstream ORFs did not show homology to other sequences in GenBank. At 231 kDa, the predicted mass of the product of ORF1 is among the largest for members of the *Tymovirales*. Even though members of the *Tymovirales* assemble flexuous rod-shaped or spherical particles from coat proteins (CPs) of 18 to 44 kDa (Adams et al., 2012c), it was not possible to identify a CP-encoding ORF based on amino acid sequence similarity. Viruses in the *Tymovirales* have genomes of 5.9 to 9.0 kb in size, which suggests that the sequence assembled for SlaMyfV1 at over 8.2 kb represents a nearly complete mycovirus genome.

3.3. Putative dsRNA mycovirus genomes

Two contigs named soybean leaf-associated botybirnavirus 1 (SlaBRV1) segment 1 (3,667 nt; KT598233) and segment 2 (2,385 nt; KT598241) each contained one large incomplete ORF. The predicted amino acid sequences of which showed significant identity (57% and 44%, respectively) with the products of *Sclerotinia sclerotiorum* botybirnavirus 1 (SsBRV1) segments 1 (6,457nt) and 2 (5,965 nt), respectively (Liu et al., 2015) (Table 1). Phylogenetic analysis showed that SlaBRV1 clustered with SsBRV1 that is a member of the recently proposed Botybirnaviridae family whose members have genomes that consist of two RNA segments of approximately 6 kb each, the larger of which encodes the viral replicase (Fig. 2) (Liu et al., 2015). Hence, the two SlaBRV1 contigs represent nearly 50% of a novel botybirnavirus genome.

Three contigs were predicted to express proteins related to members of the *Partitiviridae*. The genomes of members of the *Partitiviridae* consist of two linear dsRNA segments of about 1.2 to 2.4 kb each, which contain one ORF each (Nibert et al., 2014). One contig named soybean leaf-associated alphapartitivirus 1 segment 1 (1,924 nt; KT598242) was most similar to the replicase of *Rhizoctonia fumigata* partitivirus (Table 1). Contigs named soybean leaf-associated alphapartitivirus 2 segment 1 (1,871 nt; KT598245) and segment 2 (1,639 nt; KT598244) were most similar to the replicase and putative CP of *Sclerotinia sclerotiorum* partitivirus S (1,874; NC_013014.1 and 1,852 nt; NC_013015.1), respectively.

Two contigs were identified with similarity to members of the *Totiviridae*, whose genomes consist of one linear dsRNA segment of 4.6 to 7.0 kb that contain two large, usually overlapping, ORFs (Wickner et al., 2012). The 5'-proximal ORF (ORF1) expresses the putative

CP and the 3'-proximal ORF (ORF2) expresses the viral replicase. The predicted amino acid sequences of ORFs 1 and 2 of one contig (5,285 nt; KT598230) were greater than 70% identical to Coniothyrium minitans RNA virus (CmRV) isolated from Hubei Province in China (Cheng et al., 2003). Given the species demarcation criteria for members of the *Totiviridae*, the contig represented an Illinois isolate of CmRV. The host of the virus was not determined. The Chinese isolate of CmRV has a 4,975 bp genome compared to 5,285 bp for the Illinois isolate. Most of the difference in genome size resulted from the insertion of 330 bp into the 5' noncoding region of the virus genome, which did not increase the size of the product of ORF1. The second contig, soybean leaf-associated dsRNA virus 1 (2,827 nt; KT598236) contained a single large ORF that was predicted to encode an 841-amino acid protein that was 46% identical to the replicase of Botryosphaeria dothidea victorivirus 1 (5,322 bp) (Zhai et al., 2015) (Fig. 2). Hence, the contig likely represented the 3' half of a novel victorivirus genome.

3.4. Putative endornavirus genome

The predicted amino acid sequence of one contig, named soybean leaf-associated endornavirus 1 (SlaEV1; 4083 nt; KT598232), contained a conserved RdRP_2 (PF00978) domain and was 66% identical to the corresponding region of the polyprotein of Phytophthora endornavirus 1 (PEV1; 13,883 nt) (Hacker et al., 2005) (Table 1). As reported by Roossinck et al. (2011), phylogenetic analysis of available endornaviruses using the conserved region grouped sequences into two clades, one with larger genomes (at least 12.5 kb) and a second with smaller genomes (less than 11.0 kb) (Fig. 3). The smaller-genome endornaviruses included four viruses recently described from the three fungal species *Alternaria brassicicola* (Shang et al., 2015),

Sclerotinia sclerotiorum (NC_021706) (Khalifa and Pearson, 2014b) and *Tuber aestivum* (Stielow et al., 2011). The SlaEV1 predicted amino acid sequence grouped with the large-genome fungal endornaviruses including viruses from *Helicobasidium mompa*, *Rhizoctonia cerealis* and *R. soani* (Fig. 3). Based on the genome sizes of the large-genome endornaviruses (12.5 – 17.5 kb), we estimate that the SlaEV1 contig represents approximately 23 to 33% coverage of the full SlaEV1 genome.

3.5. Putative mitoviruses and ourmiaviruses

Five contigs (soybean leaf-associated mitovirus 1; SlaMV1, 2,519 nt, KT598238; SlaMV2, 2,477 nt, KT598239; SlaMV3, 2,471 nt, KT598240; SlaMV4, 1,611 nt, KT598246; SlaMV5, 2,665 nt, KT598237) had similarity to members of the *Mitovirus* genus in the family *Narnaviridae* and two (soybean leaf-associated ourmiavirus 1 and 2; SlaOurV1; 3,010 nt; KT598235; SlaOurV2; 1,589 nt; KT598247) had affinity to viruses in the *Ourmiavirus* genus (Table 1). Four of the five contigs (SlaMV1, SlaMV2, SlaMV3, and SlaMV5) likely represented nearly complete mitovirus genome sequences. Mitoviruses replicate within mitochondria and have genomes ranging from 2.3 to 3.6 kb that contain a single ORF encoding the viral replicase (Hillman and Esteban, 2012). Ourmiaviruses have been reported from plants and have genomes consisting of three ssRNA segments with sizes of 2.8 kb, 1.1 kb and 1.0 kb, the largest RNA encoding the viral replicase, which is phylogenetically related to those of the *Narnaviridae* (Rastgou et al., 2012). Because mitoviruses replicate in mitochondria, they have mitochondrial codon usage, while ourmiaviruses, which are thought to replicate in the cytoplasm (Crivelli et al., 2011), have eukaryotic nuclear codon usage. The amino acid sequences of the mitovirus-like

contigs predicted using a mitochondrial translation table clustered with mitoviruses of filamentous fungi, whereas amino acid sequences of the two ourmiavirus-like contigs predicted using the universal eukaryotic translation table clustered with *Saccharomyces* 20S RNA narnavirus (2,514 nt) (Rodriguez-Cousino et al., 1998) and plant ourmiaviruses (Fig. 4).

3.6. Putative negative-stranded RNA viruses

Four contigs were predicted to encode proteins with similarity to negative-stranded RNA viruses and were named soybean leaf-associated negative-stranded RNA viruses 1 – 4 (SlaNSRV-1, 9,041 nt, KT598225; SlaNSRV-2, 7,321 nt, KT598227; SlaNSRV-3, 6,218 nt, KT598228 and SlaNSRV-4, 5,317 nt, KT598229). The predicted amino acid sequences of the largest ORFs of each of the four contigs were up to 55% identical to the L replicase of *Sclerotinia sclerotiorum* negative-stranded RNA virus 1 (SsNSRV1) (Liu et al., 2014) (Table 1) and formed a separate clade with SsNSRV1 in phylogenetic analysis (Fig. 5B). The genome of SsNSRV1 is 10,002 nt in length and predicted to contain six ORFs (ORFs 1-6). The largest contig, SlaNSRV-1, was about 1 kb shorter than the SsNSRV1 genome and contained ORFs 2-6, but lacked ORF1 (Fig. 5A). Contigs SsNSRV2, SsNSRV3 and SsNSRV4 all contained at least a partial ORF5 encoding the L replicase. The results add further support to the suggestion of Liu et al (2014) that SsNSRV1 is a member of a novel family of fungal negative-stranded RNA viruses.

3.7. Additional putative single-stranded RNA virus genomes

Three contigs, soybean leaf-associated single-stranded RNA viruses 1 – 3 (SlaSSRV1, 4,188 nt, KT598231; SlaSSRV2, 3,519 nt, KT598234 and SlaSSRV3, 1,917 nt, KT598243) showed significant similarity to unclassified mycoviruses distantly related to members of the family *Tombusviridae*. Viruses in this family have ssRNA genomes of 3.7 to 4.8 kb with two 5'-proximal ORFs where the second ORF that encodes the RdRP is expressed by readthrough of a termination codon (Rochon et al., 2012). The coat and movement proteins are expressed from downstream ORFs via two subgenomic RNAs. The predicted amino acid sequence of the largest ORF of SlaSSRV1 was 51% identical to the putative RdRp of *Sclerotinia sclerotiorum* umbra-like virus 1 (SsUIV1; 1,360 nt; AHE13862.1) (Table 1). The predicted amino acid sequences of the largest ORFs of SlaSSRV2 and SlaSSRV3 were up to 55% identical to the putative RdRP of *Magnaporthe oryzae* RNA virus (MoRV; 3,246 nt; YP_009115495.1) and formed a distinct clade with *Diaporthe ambigua* RNA virus 1 (DaRV1; 4,113 nt) (Preisig et al., 2000), MoRV and SsUIV1 (Fig. 6). When compared to the sizes of the genomes of DaRV1 and MoRV, the sequences of the two largest contigs, SlaSSRV1 and SlaSSRV2, likely represented near complete genome sequences. Like the genomes of DaRV1 and MoRV, the sequences of SlaSSRV1 and SlaSSRV2 did not contain conserved downstream ORFs, which suggests that this group of mycoviruses represents a taxonomic unit distinct from plant viruses in the family *Tombusviridae*.

3.8. Putative ssDNA mycovirus genomes

Two contigs (soybean leaf-associated gemycircularvirus 1; SlaGemV1; KT598248) were predicted to encode proteins with homology to the two products encoded by viruses with single-stranded-circular DNA genomes similar to geminiviruses that have been identified in fungi,

plants and environmental samples (Conceição-Neto et al., 2015; Du et al., 2014; Rosario et al., 2012; Sikorski et al., 2013; Yu et al., 2010) (Table 1). A new genus, Gemycircularvirus, has been proposed for this group of viruses (Rosario et al., 2012). Like many members of the family *Geminiviridae*, the genomes of gemycircularviruses contain two nonoverlapping and divergently transcribed ORFs, one encoding the CP and the second the replication-related Rep protein (Fig. 7A). Inverse PCR using DNA extracted from the 2009 leaf samples and primers within the putative SlaGemV1 CP-encoding sequence amplified a DNA fragment of approximately 2.2 kb, the 2,221 nt sequence of which contained both contigs A and B, indicating that they were derived from the same viral genome. Like most other gemycircularviruses (Rosario et al., 2012; Conceição-Neto et al., 2015), the Rep-encoding primary transcript was predicted to contain a small (154 nt) intron (Fig. 7A). The predicted amino acid sequence of SlaGemV1 CP A was 54% and 45% identical to the CPs of faecal-associated gemycircularvirus 4 (FaGemV4; 2,224 nt; YP_009109730; Sikorski et al., 2013) and *Sclerotinia sclerotiorum* hypovirulence associated DNA virus 1 (SsHaDV1; 2,166 nt; YP_003104795; Yu et al., 2010). The predicted amino acid sequence of the SlaGemV1 Rep protein was 50% and 45% identical to the Rep proteins of FaGemV4 (YP_0091097310 and SsHaDV1 (YP_003104795), respectively. The predicted amino acid sequence of SlaGemV1 CP formed a clade with SsHaDV1 and other gemycircularviruses that were derived from metagenome analyses (Fig. 7B).

4. Discussion

In this study, we identified more than 25 genome segments from at least 22 putative mycoviruses using a metatranscriptomics approach, ranging from ~1.6 kb to 9 kb for the RNA

viruses. Only one of the mycoviruses, *Coniothyrium minitans* RNA virus, had been described previously. One ssDNA virus was also identified from its CP- and Rep-encoding transcripts. All the putative virus genomes were detected in both years except for the ssDNA virus that was detected only in 2009 that shared similarity with SsHaDV1. Purified virus particles of SsHaDV1 were capable of infecting hyphae of *S. sclerotiorum*, which suggested that the virus has an extracellular phase in its infection cycle (Yu et al., 2013). SsHaDV1 also has shown promising results as a natural fungicide when applied on leaves and fungal hyphae directly (Yu et al., 2013). The identification of additional plant-associated viruses in this group may provide new agents for management of fungal plant diseases.

Studies of plant-associated mycoviruses by Al Rwahnih et al. (2011) and Roossinck et al. (2010) that sequenced dsRNA enriched extracts identified sequences similar to viruses in the families *Chrysoviridae*, *Endornaviridae*, *Hypoviridae*, *Narnaviridae*, *Partitiviridae*, *Totiviridae*, and *Tymoviridae*. The current study also detected sequences similar to negative-stranded RNA and ssDNA viruses. The unbiased nature of total RNA transcriptome analysis allows detection of viruses of different genome types regardless of levels of dsRNA accumulation. Many of the contigs identified here represented nearly complete viral genome sequences, including the sequences of putative mycoviruses similar to members of the *Mycoviridae*, *Narnaviridae*, and *Partitiviridae* and order *Mononegavirales* – in some cases, from less than 1000 sequence reads (Table 1). The implementation of methods like those reported by Machida and Lin (2014) to determine 5' end sequences using Illumina sequencers could facilitate determination of full complete genomes in a high-throughput manner.

Analyses of viral metagenomes of environmental samples show that the field of virology has discovered less than 1% of the existing viral diversity, and the rate of discoveries by

metagenomics is higher than all other methods combined (Mokili et al., 2012). A limitation of the use of high throughput sequencing for viral metagenomics is that only viral sequences containing conserved domains can be retrieved (Mokili et al., 2012; Roossinck, 2015). For example, ourmiaviruses in plants usually have three genome segments, but using our approach, we found only the genome segment that encoded the highly conserved RdRp. RNAs 2 and 3 either do not share sufficient sequence similarity with previously described viruses to be detected or do not exist in fungal ourmiaviruses. The metatranscriptomes represented by soybean leaf total RNA contained relatively few sequence reads similar to mycoviruses. While techniques that enriched for mycoviral nucleic acids may have increased the depth of coverage and therefore the length of a subset of the contigs, the present study provides an unbiased survey of mycoviruses associated with soybean leaves that can be further exploited as natural genetic resources.

Sequences that appeared to have been derived from fungal viruses in the families *Chrysoviridae*, *Endornaviridae*, *Partitiviridae* and *Totiviridae* have been found to be integrated into the nuclear genomes of fungi and plants (Aiewsakun and Katzourakis 2015). In addition, the draft genome sequence of the grapevine dieback fungus *Eutypa lata* contained a nearly full-length copy of a viral genome similar to plant-infecting members of the *Virgaviridae* (Blanco-Ulate et al., 2013). Hence, it is possible that some of the putative mycovirus genomes identified in this study could have been derived from mycoviruses integrated into either the soybean genome or the genomes of associated fungi. However, using oligonucleotide primers specific for the putative viral sequences, amplicons were produced only for the sequence related to gemycircularviruses, which have ssDNA genomes. Hence, it is unlikely that the putative mycovirus genomes were derived from endogenized viruses. The identified sequences may provide bases for synthetic virology, i.e., the design and construction of new biological entities

(Wimmer et al., 2009), that could add new features to enhance their efficacy for improving plant health.

Acknowledgements

This work was supported by a grant from the Illinois Soybean Association and the United States Department of Agriculture - Agricultural Research Service. Mention of a trademark, proprietary product, or vendor does not constitute a guarantee or warranty of the product by the USDA or the University of Illinois and does not imply its approval to the exclusion of other products or vendors that may also be suitable.

References

- Adams, M.J., Candresse, T., Hammond, J., Kreuze, J.F., Martelli, G.P., Namba, S., Pearson, M.N., Ryu, K.H., Saldarelli, P., Yoshikawa, N., 2012a. Family *Betaflexiviridae*. In: King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J. (Eds.), Virus taxonomy: classification and nomenclature of viruses: Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, Waltham, MA USA, pp. 920-941.
- Adams, M.J., Candresse, T., Hammond, J., Kreuze, J.F., Martelli, G.P., Namba, S., Pearson, M.N., Ryu, K.H., Vaira, A.M., 2012b. Family *Alphaflexiviridae*. In: King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J. (Eds.), Virus Taxonomy: Classification and Nomenclature of Viruses; Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, Waltham, MA USA, pp. 904-919.

- Adams, M.J., Kreuze, J.F., Martelli, G.P., 2012c. Order *Tymovirales*. In: King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J. (Eds.), Virus Taxonomy: Classification and Nomenclature of Viruses; Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, Waltham, MA USA, pp. 901-902.
- Adams, M.J., Kreuze, J.F., Pearson, M.N., 2012d. Family *Gammapflexiviridae*. In: King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J. (Eds.), Virus Taxonomy: Classification and Nomenclature of Viruses; Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, Waltham, MA USA, pp. 942-943.
- Aiewsakun, P., Katzourakis, A., 2015. Endogenous viruses: Connecting recent and ancient viral evolution. *Virology* 479, 26-37.
- Al Rwahnih, M., Daubert, S., Urbez-Torres, J.R., Cordero, F., Rowhani, A., 2011. Deep sequencing evidence from single grapevine plants reveals a virome dominated by mycoviruses. *Arch. Virol.* 156, 397-403.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. *J. Mol. Biol.* 215, 403-410.
- Anagnostakis, S.L., 1982. Biological control of chestnut blight. *Science* 215, 466-471.
- Blackwell, M., 2011. The Fungi: 1, 2, 3 ... 5.1 million species? *Am. J. Bot.* 98, 426-438.
- Blanco-Ulate, B., Rolshausen, P.E., Cantu, D., 2013. Draft genome sequence of the grapevine dieback fungus *Eutypa lata* UCR-EL1. *Genome Announc.* 1, e00228-13.
- Brasier, C.M., 1983. A cytoplasmically transmitted disease of *Ceratocystis ulmi*. *Nature* 305, 220-223.
- Buchfink, B., Xie, C., Huson, D.H., 2015. Fast and sensitive protein alignment using DIAMOND. *Nat. Meth.* 12, 59-60.

- Chandler, J.A., Liu, R.M., Bennett, S.N., 2015. RNA shotgun metagenomic sequencing of northern California (USA) mosquitoes uncovers viruses, bacteria, and fungi. *Front. Microbiol.* 6, 185.
- Cheng, J., Jiang, D., Fu, Y., Li, G., Peng, Y., Ghabrial, S.A., 2003. Molecular characterization of a dsRNA totivirus infecting the sclerotial parasite *Coniothyrium minitans*. *Virus Res.* 93, 41-50.
- Chiba, S., Salaipeh, L., Lin, Y.-H., Sasaki, A., Kanematsu, S., Suzuki, N., 2009. A novel bipartite double-stranded RNA mycovirus from the white root rot fungus *Rosellinia necatrix*: molecular and biological characterization, taxonomic considerations, and potential for biological control. *J. Virol.* 83, 12801-12812.
- Coetzee, B., Freeborough, M.-J., Maree, H.J., Celton, J.-M., Rees, D.J.G., Burger, J.T., 2010. Deep sequencing analysis of viruses infecting grapevines: virome of a vineyard. *Virology* 400, 157-163.
- Conceição-Neto, N., Zeller, M., Heylen, E., Lefrère, H., Mesquita, J.R., Matthijnsens, J., 2015. Fecal virome analysis of three carnivores reveals a novel nodavirus and multiple gemycircularviruses. *Virol. J.* 12, 79.
- Crivelli, G., Ciuffo, M., Genre, A., Masenga, V., Turina, M., 2011. Reverse genetic analysis of ourmiaviruses reveals the nucleolar localization of the coat protein in *Nicotiana benthamiana* and unusual requirements for virion formation. *J. Virol.* 85, 5091-5104.
- Du, Z., Tang, Y., Zhang, S., She, X., Lan, G., Varsani, A., He, Z., 2014. Identification and molecular characterization of a single-stranded circular DNA virus with similarities to *Sclerotinia sclerotiorum* hypovirulence-associated DNA virus 1. *Arch. Virol.* 159, 1527-1531.

- Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucl. Acids Res.* 32, 1792-1797.
- Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460-2461.
- Fehr, W.R., Caviness, C. E., Burmood, D. T., Pennington, J. S., 1971. Stage of development descriptions for soybeans, *Glycine Max* (L.) Merrill. *Crop Sci.* 11, 929-931.
- Ghabrial, S.A., Castón, J.R., Jiang, D., Nibert, M.L., Suzuki, N., 2015. 50-plus years of fungal viruses. *Virology* 479, 356-368.
- Ghabrial, S.A., Soldevila, A.I., Havens, W.M., 2002. Molecular genetics of the viruses infecting the plant pathogenic fungus *Helminthosporium victoriae*. In: Tavantzis, S. (Ed.), *Molecular biology of double-stranded RNA: concepts and applications in agriculture, forestry and medicine*. CRC Press, Boca Raton, pp. 213–236.
- Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., Di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644-652.
- Hacker, C.V., Brasier, C.M., Buck, K.W., 2005. A double-stranded RNA from a *Phytophthora* species is related to the plant endornaviruses and contains a putative UDP glycosyltransferase gene. *J. Gen. Virol.* 86, 1561-1570.
- Hale, I.L., Broders, K., Iriarte, G., 2014. A Vavilovian approach to discovering crop-associated microbes with potential to enhance plant immunity. *Front. Plant Sci.* 5, 7.

- Hillman, B.I., Esteban, R., 2012. Family *Narnaviridae*. In: King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J. (Eds.), *Virus Taxonomy: Classification and Nomenclature of Viruses; Ninth Report of the International Committee on Taxonomy of Viruses*. Elsevier Academic Press, Waltham, MA USA, pp. 1055-1060.
- Hobbs, H.A., Herman, T.K., Slaminko, T.L., Wang, Y., Nguyen, B.T., McCoppin, N.K., Domier, L.L., Hartman, G.L., 2010. Occurrences of soybean viruses, fungal diseases, and pests in Illinois soybean rust sentinel plots. *Plant Health Prog.* doi:10.1094/PHP-2010-0827-01-BR.
- Hong, Y.G., Dover, S.L., Cole, T.E., Brasier, C.M., Buck, K.W., 1999. Multiple mitochondrial viruses in an isolate of the Dutch elm disease fungus *Ophiostoma novo ulmi*. *Virology* 258, 118-127.
- Howitt, R.L.J., Beever, R.E., Pearson, M.N., Forster, R.L.S., 2001. Genome characterization of Botrytis virus F, a flexuous rod-shaped mycovirus resembling plant 'potex-like' viruses. *J. Gen. Virol.* 82, 67-78.
- Hu, Z.J., Wu, S.S., Cheng, J.S., Fu, Y.P., Jiang, D.H., Xie, J.T., 2014. Molecular characterization of two positive-strand RNA viruses co-infecting a hypovirulent strain of *Sclerotinia sclerotiorum*. *Virology* 464, 450-459.
- Huson, D.H., Mitra, S., Ruscheweyh, H.J., Weber, N., Schuster, S.C., 2011. Integrative analysis of environmental sequences using MEGAN4. *Genome Res.* 21, 1552-1560.
- Khalifa, M.E., Pearson, M.N., 2014a. Characterisation of a novel hypovirus from *Sclerotinia sclerotiorum* potentially representing a new genus within the *Hypoviridae*. *Virology* 464–465, 441-449.

- Khalifa, M.E., Pearson, M.N., 2014b. Molecular characterisation of an endornavirus infecting the phytopathogen *Sclerotinia sclerotiorum*. *Virus Res.* 189, 303-309.
- Kondo, H., Chiba, S., Toyoda, K., Suzuki, N., 2013a. Evidence for negative-strand RNA virus infection in fungi. *Virology* 435, 201-209.
- Kondo, H., Kanematsu, S., Suzuki, N., 2013b. Viruses of the white root rot fungus, *Rosellinia necatrix*. In: Ghabrial, S.A. (Ed.), *Advances in Virus Research*, Vol 86: Mycoviruses. Vol. 86, pp. 177-214.
- Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Meth.* 9, 357-359.
- Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25.
- Leonard, K.J., Szabo, L.J., 2005. Stem rust of small grains and grasses caused by *Puccinia graminis*. *Mol. Plant Pathol.* 6, 99-111.
- Liu, L., Xie, J., Cheng, J., Fu, Y., Li, G., Yi, X., Jiang, D., 2014. Fungal negative-stranded RNA virus that is related to bornaviruses and nyaviruses. *Proc. Natl. Acad. Sci. USA* 111, 12205-12210.
- Liu, L.J., Wang, Q.H., Cheng, J.S., Fu, Y.P., Jiang, D.H., Xie, J.T., 2015. Molecular characterization of a bipartite double-stranded RNA virus and its satellite like RNA co-infecting the phytopathogenic fungus *Sclerotinia sclerotiorum*. *Frontiers in Microbiology* 6, 406.
- Machida, R.J., Lin, Y.Y., 2014. Four methods of preparing mRNA 5' end libraries using the illumina sequencing platform. *PLoS One* 9, e101812.

- Márquez, L.M., Redman, R.S., Rodriguez, R.J., Roossinck, M.J., 2007. A virus in a fungus in a plant: three-way symbiosis required for thermal tolerance. *Science* 315, 513-515.
- Marvelli, R.A., Hobbs, H.A., Li, S., McCoppin, N.K., Domier, L.L., Hartman, G.L., Eastburn, D.M., 2014. Identification of novel double-stranded RNA mycoviruses of *Fusarium virguliforme* and evidence of their effects on virulence. *Arch. Virol.* 159, 349-352.
- Marzano, S.Y.L., Hobbs, H.A., Nelson, B.D., Hartman, G.L., Eastburn, D.M., McCoppin, N.K., Domier, L.L., 2015. Transfection of *Sclerotinia sclerotiorum* with in vitro transcripts of a naturally occurring interspecific recombinant of *Sclerotinia sclerotiorum* hypovirus 2 significantly reduces virulence of the fungus. *J. Virol.* 89, 5060-5071.
- McSpadden-Gardener, B., Fravel, D., 2002. Biological control of plant pathogens: Research, commercialization, and application in the USA. *Plant Health Prog.* doi:10.1094/PHP-2002-0510-01-RV.
- Mokili, J.L., Rohwer, F., Dutilh, B.E., 2012. Metagenomics and future perspectives in virus discovery. *Curr. Opin. Virol.* 2, 63-77.
- Nibert, M.L., Ghabrial, S.A., Maiss, E., Lesker, T., Vainio, E.J., Jiang, D.H., Suzuki, N., 2014. Taxonomic reorganization of family *Partitiviridae* and other recent progress in partitivirus research. *Virus Res.* 188, 128-141.
- Nuss, D.L., 1992. Biological control of chestnut blight - an example of virus-mediated attenuation of fungal pathogenesis. *Microbiol. Rev.* 56, 561-576.
- Perez-Garcia, A., Romero, D., Fernandez-Ortuno, D., Lopez-Ruiz, F., De Vicente, A., Tores, J.A., 2009. The powdery mildew fungus *Podosphaera fusca* (synonym *Podosphaera xanthii*), a constant threat to cucurbits. *Molec. Plant Pathol.* 10, 153-160.

- Preisig, O., Moleleki, N., Smit, W.A., Wingfield, B.D., Wingfield, M.J., 2000. A novel RNA mycovirus in a hypovirulent isolate of the plant pathogen *Diaporthe ambigua*. J. Gen. Virol. 81, 3107-3114.
- Rastgou, M., Habibi, M.K., Izadpanah, K., Masenga, V., Milne, R.G., Wolf, Y.I., Koonin, E.V., Turina, M., 2009. Molecular characterization of the plant virus genus *Ourmiavirus* and evidence of inter-kingdom reassortment of viral genome segments as its possible route of origin. J. Gen. Virol. 90, 2525-2535.
- Rastgou, M., Turina, M., Milne, R.G., 2012. Genus *Ourmiavirus*. In: King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J. (Eds.), Virus Taxonomy: Classification and Nomenclature of Viruses; Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, Waltham, MA USA, pp. 1177-1180.
- Rochon, D., Lommel, S., Martelli, G.P., Rubino, L., Russo, M., 2012. Family *Tombusviridae*. In: King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J. (Eds.), Virus Taxonomy: Classification and Nomenclature of Viruses; Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, Waltham, MA USA, pp. 1111-1138.
- Rodriguez-Cousino, N., Solorzano, A., Fujimura, T., Esteban, R., 1998. Yeast positive-stranded virus-like RNA replicons - 20 S and 23 S RNA terminal nucleotide sequences and 3' end secondary structures resemble those of RNA coliphages. J. Biol. Chem. 273, 20363-20371.
- Rodriguez, R.J., White, J.F., Arnold, A.E., Redman, R.S., 2009. Fungal endophytes: diversity and functional roles. New Phytol. 182, 314-330.
- Roossinck, M.J., 2015. Metagenomics of plant and fungal viruses reveals an abundance of persistent lifestyles. Front. Microbiol. 5, 767.

- Roossinck, M.J., Sabanadzovic, S., Okada, R., Valverde, R.A., 2011. The remarkable evolutionary history of endornaviruses. *J. Gen. Virol.* 92, 2674-2678.
- Roossinck, M.J., Saha, P., Wiley, G.B., Quan, J., White, J.D., Lai, H., Chavarria, F., Shen, G., Roe, B.A., 2010. Ecogenomics: using massively parallel pyrosequencing to understand virus ecology. *Molec. Ecol.* 19, 81-88.
- Rosario, K., Dayaram, A., Marinov, M., Ware, J., Kraberger, S., Stainton, D., Breitbart, M., Varsani, A., 2012. Diverse circular ssDNA viruses discovered in dragonflies (Odonata: Epiprocta). *J. Gen. Virol.* 93, 2668-2681.
- Ruby, J.G., Bellare, P., DeRisi, J.L., 2013. PRICE: software for the targeted assembly of components of (meta) genomic sequence data. *G3* 3, 865-880.
- Schoebel, C.N., Zoller, S., Rigling, D., 2014. Detection and genetic characterisation of a novel mycovirus in *Hymenoscyphus fraxineus*, the causal agent of ash dieback. *Infect. Genet. and Evol.* 28, 78-86.
- Shang, H.H., Zhong, J., Zhang, R.J., Chen, C.Y., Gao, B.D., Zhu, H.J., 2015. Genome sequence of a novel endornavirus from the phytopathogenic fungus *Alternaria brassicicola*. *Arch. Virol.* 160, 1827-1830.
- Sikorski, A., Massaro, M., Kraberger, S., Young, L.M., Smalley, D., Martin, D.P., Varsani, A., 2013. Novel myco-like DNA viruses discovered in the faecal matter of various animals. *Virus Res.* 177, 209-216.
- Stielow, B., Klenk, H.P., Menzel, W., 2011. Complete genome sequence of the first endornavirus from the ascocarp of the ectomycorrhizal fungus *Tuber aestivum* Vittad. *Arch. Virol.* 156, 343-345.

- Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725-2729.
- Vainio, E.J., Jurvansuu, J., Streng, J., Rajamaki, M.L., Hantula, J., Valkonen, J.P.T., 2015. Diagnosis and discovery of fungal viruses using deep sequencing of small RNAs. *J. Gen. Virol.* 96, 714-725.
- Wickner, R.B., Ghabrial, S.A., Nibert, M.L., Patterson, J.L., Wang, C.C., 2012. Family *Totiviridae*. In: King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J. (Eds.), *Virus Taxonomy: Classification and Nomenclature of Viruses; Ninth Report of the International Committee on Taxonomy of Viruses*. Elsevier Academic Press, Waltham, MA USA, pp. 639-650.
- Wimmer, E., Mueller, S., Tumpey, T.M., Taubenberger, J.K., 2009. Synthetic viruses: a new opportunity to understand and prevent viral disease. *Nat. Biotechnol.* 27, 1163-1172.
- Woo, S.L., Ruocco, M., Vinale, F., Nigro, M., Marra, R., Lombardi, N., Pascale, A., Lanzuise, S., Manganiello, G., Lorito, M., 2014. *Trichoderma*-based products and their widespread use in agriculture. *Open Myco. J.* 8, 71–126.
- Yu, X., Li, B., Fu, Y., Jiang, D., Ghabrial, S.A., Li, G., Peng, Y., Xie, J., Cheng, J., Huang, J., Yi, X., 2010. A geminivirus-related DNA mycovirus that confers hypovirulence to a plant pathogenic fungus. *Proc. Natl. Acad. Sci. USA* 107, 8387-8392.
- Yu, X., Li, B., Fu, Y., Xie, J., Cheng, J., Ghabrial, S.A., Li, G., Yi, X., Jiang, D., 2013. Extracellular transmission of a DNA mycovirus and its use as a natural fungicide. *Proc. Natl. Acad. Sci. USA* 110, 1452-1457.

Zhai, L.F., Hong, N., Zhang, M.X., Wang, G.P., 2015. Complete dsRNA sequence of a novel victorivirus isolated from the pear stem wart fungus *Botryosphaeria dothidea*. Arch. Virol. 160, 613-616.

Figure Legends

Fig. 1. Genome organization and phylogenetic relationships of soybean leaf-associated mycoflexivirus 1 (SlaMyfV1). (A) Comparisons of the organization of the SlaMyfV1 with that of Botrytis virus X (BotVX). Open reading frames (ORFs) are shown as boxes and staggered to indicate the reading frame. (B) Maximum-likelihood tree depicting the relationships of the predicted amino acid sequence of RNA-dependent RNA polymerases (RdRps) of SlaMyfV1 and other confirmed and proposed members of the *Tymovirales*. Predicted RdRp amino acid sequences were aligned with MUSCLE, and trees were inferred using MEGA6. Branch lengths are scaled to the expected underlying number of amino acid substitutions per site. Bootstrap percentages greater than 50% are shown. The Botrytis virus X RdRp amino acid sequence was used as an outgroup to root the tree.

Fig. 2. Maximum-likelihood tree depicting the relationships of the predicted amino acid sequence of RNA-dependent RNA polymerase (RdRp) of the soybean leaf-associated double-stranded RNA viruses and other confirmed and proposed members of the genera Botybirnavirus, *Alphapartitivirus*, *Totivirus* and *Victorivirus*. Predicted RdRp amino acid sequences were aligned and phylogenetic trees were constructed as described in Fig. 1.

Fig. 3. Maximum-likelihood tree depicting the relationships of the predicted amino acid sequence of RNA-dependent RNA polymerase (RdRp) of the soybean leaf-associated endornavirus with other confirmed and proposed members of the *Endornaviridae*. Predicted RdRp amino acid sequences were aligned and phylogenetic trees were constructed as described

in Fig. 1. The lengths of genomic sequences are indicated in parenthetically. Asterisks indicate partial sequences.

Fig. 4. Maximum-likelihood tree depicting the relationships of the predicted amino acid sequence of RNA-dependent RNA polymerase (RdRp) of the soybean leaf-associated mitoviruses and ourmiaviruses with other confirmed and proposed members of the family *Narnaviridae* and genus *Ourmiavirus*. Predicted RdRp amino acid sequences were aligned and phylogenetic trees were constructed as described in Fig. 1. The cucumber necrosis virus RdRp amino acid sequence was used as an outgroup to root the tree.

Fig. 5. Genome organization and phylogenetic relationships of soybean leaf-associated negative-stranded RNA viruses 1 – 5 (SlaNRV1-5). (A) Comparisons of the organizations of SlaNRV1-5 with that of *Sclerotinia sclerotiorum* negative-stranded RNA virus 1 (SsNRV1). (B) Maximum likelihood tree depicting the relationships of the predicted amino acid sequence of RNA-dependent RNA polymerase (RdRp) of the soybean leaf-associated negative-stranded RNA viruses and other confirmed and proposed members of the *Mononegavirales*. Predicted RdRp amino acid sequences were aligned and phylogenetic trees were constructed as described in Fig. 1.

Fig. 6. Genome organization and phylogenetic relationships of soybean leaf-associated single-stranded RNA viruses 1 and 2 (SlaSSRV1 and SlaSSRV2). (A) Comparisons of the organizations of the SlaSSRV1 and SlaSSRV2 with that of *Diaporthe ambigua* RNA virus 1 (DaRV1) and *Magnaporthe oryzae* RNA virus (MoRV). (B) Maximum likelihood tree depicting the relationships of the predicted amino acid sequence of RNA-dependent RNA polymerase

(RdRp) of the soybean leaf-associated ssRNA viruses and other confirmed and proposed members of the family *Tombusviridae* and genus *Umbravirus*. Predicted RdRp amino acid sequences were aligned and phylogenetic trees were constructed as described in Fig. 1.

Fig. 7. Genome organization and phylogenetic relationships of soybean leaf-associated gemycircularvirus 1. (A) Gene organization of the soybean leaf-associated gemycircularvirus 1 indicating the positions of the contigs A and B that showed amino acid sequence similarity to putative coat protein (CP) and replicase-related proteins of gemycircularviruses. (B) Maximum-likelihood tree depicting the relationships of the predicted amino acid sequence of CP of the soybean leaf-associated gemycircularvirus 1 and other confirmed and proposed members of the gemycircularvirus genus. Predicted RdRp amino acid sequences were aligned and phylogenetic trees were constructed as described in Fig. 1.

Table 1.

Mycovirus-like genomes recovered from field-grown soybean leaf samples.

Contig name	GenBank Accession	Read counts		Contig length (nt)	Ident. (%)	Putative function [most similar virus]
		2008	2009			
Soybean leaf-associated mycoflexivirus 1	KT598226	17,856	618	8,327	26	RdRP ^a [Botrytis virus F]
Soybean leaf -associated botybirnavirus 1 RNA1	KT598233	236	12	3,667	36	hypothetical protein [Sclerotinia sclerotiorum botybirnavirus 1]
Soybean leaf -associated botybirnavirus 1 RNA2	KT598241	80	4	2,385	57	cap-pol fusion protein [Sclerotinia sclerotiorum botybirnavirus 1]
Soybean leaf -associated partitivirus 1 RNA1	KT598242	282	6	1,924	44	RdRp [Rhizoctonia fumigata partitivirus]
Soybean leaf -associated partitivirus 2 RNA1	KT598245	1,190	34	1,871	54	RdRp [Sclerotinia sclerotiorum partitivirus S]
Soybean leaf -associated partitivirus 2 RNA2	KT598244	230	2	1,639	31	CP ^b [Sclerotinia sclerotiorum partitivirus S]
Soybean leaf -associated dsRNA virus 1	KT598236	474	12	2,827	46	RdRp [Botryosphaeria dothidea victorivirus 1]
Coniothyrium minitans RNA virus-Illinois	KT598230	1,008	10	5,285	70	RdRp [Coniothyrium minitans RNA virus]
Soybean leaf -associated endornavirus 1	KT598232	184	122	4,083	44	RdRp [Phytophthora endornavirus 1]
Soybean leaf -associated mitovirus 1	KT598238	1,544	38	2,519	41	RdRp [Sclerotinia sclerotiorum mitovirus 7]
Soybean leaf -associated mitovirus 2	KT598239	374	6	2,477	64	RdRp [Alternaria brassicicola mitovirus]
Soybean leaf -associated mitovirus 3	KT598240	1,440	920	2,471	38	RdRp [Mitovirus AEF-2013]
Soybean leaf -associated mitovirus 4	KT598246	224	4	1,611	55	RdRp [Alternaria brassicicola mitovirus]
Soybean leaf -associated mitovirus 5	KT598237	2,330	54	2,665	37	RdRp [Thanatephorus cucumeris]
Soybean leaf -associated ourmiavirus 1	KT598235	4,446	318	3,010	34	RdRp [Ourmia melon virus]
Soybean leaf -associated ourmiavirus 2	KT598247	174	20	1,589	30	RdRp [Ourmia melon virus]
Soybean leaf -associated negative-stranded RNA virus 1	KT598225	896	36	9,041	55	RdRp [Sclerotinia sclerotiorum negative-stranded RNA virus 1]
Soybean leaf -associated negative-stranded RNA virus 2	KT598227	716	22	7,321	54	RdRp [Sclerotinia sclerotiorum negative-stranded RNA virus 1]
Soybean leaf -associated negative-stranded RNA	KT598228	926	54	6,218	30	RdRp [Sclerotinia sclerotiorum

virus 3							negative-stranded RNA virus 1]
Soybean leaf -associated negative-stranded RNA virus 4	KT598229	632	24	5,317	29		RdRp [Sclerotinia sclerotiorum negative-stranded RNA virus 1]
Soybean leaf -associated ssRNA virus 1	KT598231	2,486	18	4,188	51		SsULV1gp2 [Sclerotinia sclerotiorum umbra-like virus 1]
Soybean leaf -associated ssRNA virus 2	KT598234	7,532	190	3,519	54		RdRp [Magnaporthe oryzae RNA virus]
Soybean leaf -associated ssRNA virus 3	KT598243	362	12	1,917	55		RdRp [Magnaporthe oryzae RNA virus]
Soybean leaf -associated ssDNA virus 1 CP contig A	KT598248	0	104	665	47		CP [Cassava associated circular DNA virus]
Soybean leaf -associated ssDNA virus 1 rep contig B	KT598248	0	176	264	55		Rep ^c [Sclerotinia sclerotiorum hypovirulence associated DNA virus 1]
<hr/>							
^a RNA-dependent RNA polymerase							
^b Coat protein							
^c Replication-associated protein							

Figure 1

A

B

Figure 2

Figure 3

Figure 4

Figure 5

A

B

Figure 6

A

B

Figure 7

A

B